

# Presenting and processing information in background noise: A combined speaker–listener perspective

Annelies Bockstael<sup>a)</sup>

*École d'orthophonie et d'audiologie, Université de Montréal, Montreal, Québec, H3N 1X7, Canada*

Laurie Samyn and Paul Corthals

*Department of Speech, Language and Hearing Sciences, Ghent University, Ghent, 9000, Belgium*

Dick Botteldooren

*Department of Information Technology, Ghent University, Zwijnaarde, 9052, Belgium*

(Received 14 June 2017; revised 2 November 2017; accepted 26 December 2017; published online 16 January 2018)

Transferring information orally in background noise is challenging, for both speaker and listener. Successful transfer depends on complex interaction between characteristics related to listener, speaker, task, background noise, and context. To fully assess the underlying real-life mechanisms, experimental design has to mimic this complex reality. In the current study, the effects of different types of background noise have been studied in an ecologically valid test design. Documentary-style information had to be presented by the speaker and simultaneously acquired by the listener in four conditions: quiet, unintelligible multitalker babble, fluctuating city street noise, and little varying highway noise. For both speaker and listener, the primary task was to focus on the content that had to be transferred. In addition, for the speakers, the occurrence of hesitation phenomena was assessed. The listener had to perform an additional secondary task to address listening effort. For the listener the condition with the most eventful background noise, i.e., fluctuating city street noise, appeared to be the most difficult with markedly longer duration of the secondary task. In the same fluctuating background noise, speech appeared to be less disfluent, suggesting a higher level of concentration from the speaker's side. © 2018 Acoustical Society of America.

<https://doi.org/10.1121/1.5020799>

[JFL]

Pages: 210–218

## I. INTRODUCTION

Epidemiological data vastly show that excessive background noise has a negative effect on learning and cognitive performance.<sup>1</sup> Laboratory experiments confirm that background noise hampers processing of oral information.<sup>2</sup> This is not only a matter of impaired intelligibility. Even in listening conditions where the signal-to-noise ratio assures good speech intelligibility, background noise still has an adverse effect on information acquisition.<sup>2</sup>

Not only information processing (task performance) as such, but also the effort needed to achieve a certain level of performance might be affected by background noise. In this, listening effort can be defined as mental effort to overcome obstacles in goal pursuit when carrying out a task.<sup>3,4</sup> When speech is harder to understand, more cognitive resources are needed for speech processing.<sup>4</sup> Effort is clearly a different aspect of (listening) tasks than final performance. No direct relationship is found between performance on a listening task and the associated effort.<sup>5</sup>

Listening effort can be assessed in different ways, via behavioral testing, electrophysiological measurements, and self-reporting.<sup>3</sup> Behaviorally, most common measures are those that index working memory, attention, and processing speed.<sup>3</sup> In the classical dual-task paradigm, focus and

attention are assessed by imposing two different tasks.<sup>5</sup> One task is the primary task, i.e., the listening task, and the listener is instructed to optimize performance for this task. The secondary task is used as a competing task. Theoretically, it is assumed that if the required resources to perform both tasks exceed a person's available cognitive and attentional resources, the person's processing system will prioritize the primary task to the detriment of the secondary task.<sup>5</sup>

Listening effort is mostly measured with a concurrent dual-task paradigm where the primary and secondary task have to be performed at the same time—instead of sequentially.<sup>5</sup> The advantage of a concurrent design is higher ecological validity, as many real-life speech processing situations require multitasking.<sup>5</sup> In teaching environments, for instance, students are not only listening to the teacher, but also taking notes, reading the blackboard, perhaps communicating with their neighbors.

Dual-task studies on listening effort mostly use word or sentence recognition,<sup>5</sup> but the paradigm has also been applied to documentary-style passages with comprehension—instead of mere recognition—as primary task.<sup>6,7</sup> For the secondary task, for instance, a visual or tactile task, performance and response time can be measured. So far, no particular type of secondary task has been shown to be especially suitable to assess listening effort.<sup>5</sup>

As for background noise, different characteristics appear to affect presentation, perception and processing of

<sup>a)</sup>Electronic mail: [annelies.bockstael@umontreal.ca](mailto:annelies.bockstael@umontreal.ca)

information differently. A first clear distinction is made between speech and non-speech background noise. For background noise with competing talkers, especially intelligible background, speech appears to be cumbersome.<sup>4</sup> The strong distracting effect of meaningful non-target speech can be explained by semantic processing,<sup>4,8</sup> but potentially also phonological similarity between target and non-target speech increases informational masking and distraction.<sup>4,9</sup> In addition, the number of competing speakers is of importance. The effect of multitalker masking appears to be the strongest with three to four competing speakers.<sup>9</sup> The effect becomes less prominent as the number of talkers increases and around 12 the masking effects saturate because the talkers in the babble are masking each other.<sup>9</sup>

For non-speech background, the effects can best be understood in terms of the acoustic characteristics of the non-target signal. The influence varies depending on the task at hand. For pure speech intelligibility tasks, less overlap in time and/or frequency domain between target speech and masker appears to be beneficial, especially for normal-hearing listeners.<sup>10</sup> Performance on short-term memory tasks appears to be particularly disturbed by distinctive temporal-spectral variations in the background noise,<sup>11</sup> especially if their occurrence is difficult to predict.<sup>12</sup> Habituation to these acoustic changes hardly occurs, and short periods of quiet cause rapid dishabituation.<sup>12</sup> When the effect of background noise is assessed for exposure lasting several hours or even days, decreased word performance, increased fatigue and motivational deficits have been found for increasing background level.<sup>13</sup>

Not only processing of information, but also oral presentation is directly influenced by the background noise. Speakers can adapt speech to the background noise to increase the probability of successful communication by changing speech level, frequency content, and temporal structure.<sup>14,15</sup> Speaking in background noise is considered to be more effortful, leaving fewer resources to focus on the content of the message that has to be presented. Background noise in classrooms is associated with more vocal symptoms for teachers and development of cognitive fatigue after work.<sup>16</sup> Disfluencies and hesitation phenomena in speech, such as filled pauses (*ums* and *ers*), corrections and repetitions might be indicative for additional cognitive load and reduced working memory capacities.<sup>17</sup>

Presenting and processing information in background noise requires the maintenance of speech intelligibility, and involves memory tasks as well as sustained attention. Predicting the final outcome is challenging because of the complex interaction between listener, speaker, task, background noise, and context. Research has found qualitatively different effects of noise exposure depending on task complexity.<sup>18,19</sup> Interestingly, especially for more complex tasks concentration<sup>18</sup> and context<sup>2</sup> offer more effective coping strategies to reduce the negative effects of background noise.

This paper addresses the effect of noise exposure when presenting and processing information in an ecologically valid design. Speaker and listener are working together in pairs to transfer documentary-style information. Three different types of realistic background noise are included:

multitalker babble, little varying highway noise, and fluctuating city street sounds. A quiet condition without additional background noise is included as reference condition.

The following four research questions will be assessed. How is the ability to focus on the content of information (to be) presented influenced by background noise, for both speaker and listener (1). How does listening effort when processing complex information varies depending on background noise conditions (2). What is the effect of background noise on speaker's fluency and hesitation phenomena (3), and how do these speech characteristics affect listening effort (4).

## II. MATERIALS AND METHODS

### A. Participants

Participant had to be fluent in Dutch (mother tongue). Exclusion criteria were reported hearing loss, severe issues with vision, problems with speech and language, and attention-related disorders. Participants were asked to have a good night's sleep prior to the experiment. The volunteers were recruited via convenience sampling. Duos who already knew each other before were preferred to facilitate the speaker-listener interaction.

In total 60 people participated, 42 female and 18 male. They were on average 21.4 yr old, the youngest participants being 18 yr old and the oldest 32. All signed the informed consent approved by the ethical committee before testing.

### B. Background noise

For the background noise, the following recordings were played back; (1) fluctuating city street noise, (2) little varying highway traffic noise, and (3) unintelligible multitalker babble. A fourth (4) condition without additional background noise has been added as reference condition. All recordings were done in real sound environments using the Ambisonics four channel three dimensional (3D) recording system.

For the city street sound, recordings were made at the corner of a one-way car lane with a bicycle lane next to it, close to a park. The fragment mainly includes individually audible car passages, bicycles passing by, and birds singing in the trees. For the highway noise, noise of dense traffic was recorded, for which no individual car passages could be recognized. For the babble noise, recordings were made at a cocktail party where about 20 people were having conversations in Dutch and in English. The recorded speech was not intelligible.

The participants were seated 2 m from the two speakers playing the sounds. The level of the background noise on play-back has been verified with a Svantek 959 sound level meter. Background noise fragments were played at 70 dB(A) (LAeq over noise fragment duration 355 s) at the participants' position. Figure 1 depicts per fragment the variation in loudness as a function of time. Loudness has been calculated in accordance to the ISO 532-1:2017 standard for calculating loudness of non-stationary sounds<sup>20</sup> and allows us to assess the perceptual strength of sounds.<sup>21</sup> Figure 1 shows

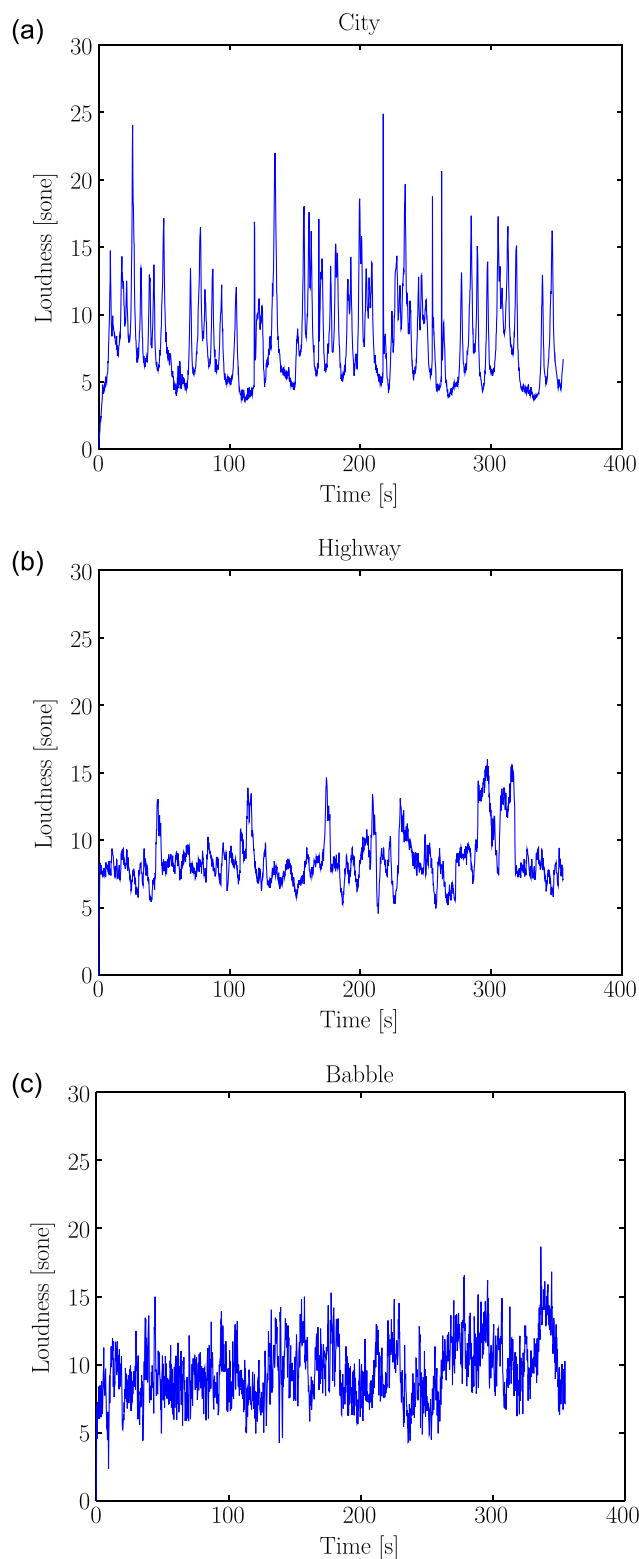


FIG. 1. (Color online) Loudness as a function of time for the three noise fragments. Noise recorded at a city street (City), highway noise (Highway), and multitalker babble (Babble).

that the city street noise clearly stands out in terms of loudness fluctuation: over the whole duration of the fragment more quiet periods of about 5 sone are altered with events reaching loudness levels of 15 sone and more. For highway noise the loudness is quite stable around 7 sone with a few distinct events (peaks up to 15 sone). Finally, the multitalker

babble shows somewhat more variation in loudness than the highway noise, but unlike the city street noise, the differences between the loudest and the most quiet loudness levels have on order of magnitude below 10 sone.

### C. Orally presented information

To mimic a teaching situation, the speaker had to read Dutch lectures on eight scientific topics. The information presented in the fragments was quite specialized but presented in layman's terms. This to ensure that the listener had little *a priori* knowledge about the topics, and was at the same time capable to understand the presented information.

The text was presented to the speaker on A4 paper format, printed in Calibri font, font size 11 with 1.15 line spacing. To increase readability, italic and bold fonts were included.

Reading of one fragment took between 3 and 5 min. Both speaker and listener were instructed to focus maximally on the content of the lectures. To stimulate this, they were told beforehand that they had to complete an exam afterwards.

The exam was the same for speaker and listener and consisted of three questions per lecture. The first two questions were open questions, one factual and one insight question. For the third question, a sentence had to be completed with a specific concept, name, or number. To grade the exam, specific keywords were defined per questions and points were given per present keywords in the participant's response.

### D. Secondary task listener

The secondary task presented to the listener was a visual task inspired by the secondary task described in Ref. 22. On a computer screen, 14 white squares were shown. Out of those 14 squares, six random squares were shortly colored green one after the other. After each sequence of six green squares, the participant was asked to indicate on a screen with only white squares which squares had become green. Both accuracy and time needed to respond were stored for each trial. During the whole lecture, five instances of the visual task were presented to the listener at random time intervals. For final analyses, the first and last trial were excluded to minimize learning effect on the results.

### E. Hesitation phenomena speaker

To score the disfluencies and hesitation phenomena of the speech afterwards, speech was recorded with a Neumann KM 88 I microphone. The microphone was placed on the table between speaker and listener using a tripod at 30 cm from the speaker's mouth. Wav-files were recorded with the microphone switch to figure-eight pattern,<sup>23</sup> meaning that especially signals coming from the direction of speaker and listener were recorded whereas omnidirectional signals, i.e., the background noise, were suppressed to ensure that recording and analyses of the speech signals would not be corrupted by the presented background noise.

With the software Praat, the total duration of disfluent speech in each speech sample was calculated. Following disfluencies were included: (1) repetitions, (2) corrections of wrongly read phonemes, syllables, words, part of sentences, and intonation patterns, (3) abnormal long or unusual pauses, (4) interjections, (5) words or word groups read in a hesitating way, and (6) subtle prolonging of fricatives and nasals.

## F. Test setup

For each test, the two participants, one speaker, one listener, were seated facing each other at 1.20 m distance. Both speaker and listener had their own computer screen, respectively, showing the text of the lecture and the visual task. The screens did not block the participants' view.

For the first four lectures, one participant was assigned speaker and the other listener. The speaker was instructed to present the information clearly, and at the same time retain as much of the information as possible. The listener was told to focus on the information and execute at the same time the visual task without losing focus on the primary listening task. No information was given on the background noise. After four lectures, the participants switched roles for the next four fragments. At the end, an exam on the content of all lectures was completed individually by each participant. For each lecture, participants were also asked to rate their interest in the subject on a five-point scale. Lectures were coupled randomly to the background noise fragments per test duo, and presentation order was also randomized across participants.

## G. Statistical analysis

Mixed model linear regression has been applied using the LME4 package of the statistical software R.<sup>24</sup> Different models have been built separately for the three outcome variables of interest: exam score, listening effort and fluency of the speaker. In all the models, the variables participant and lecture have been included as random factors. Secondary task duration and speakers' disfluencies' duration have been logarithmically transformed to account for the skewness of those variables. To investigate the relationship between the outcome and independent variables, a two-step approach has been followed.

First, for each model, the independent variables of interest have been selected by entering them as a single fixed factor in the model containing only the two random factors. The  $p$ -value for the independent variables was calculated using single term deletion.<sup>24</sup> a  $\chi^2$ -test was performed on the difference in AIC (Akaike information criterion) value of the model with and without the fixed variable under study. For the level of significance,  $\alpha = 0.05$  was used. In a similar way, two-way interaction effects between, respectively, background noise, the variable of most interest, and the other independent variables were evaluated.

Subsequently, for the three different outcome variables, a full model has been built adding together all independent variables that were statistically significant in the previous step. Significance of all independent variables in this full model was verified using again single term deletion. The final models are reported in this paper.

Statistical model assumptions have been verified by visual inspection of the residual's Q-Q plot together with Shapiro-Wilk test and Kolmogorov-Smirnov test.<sup>25</sup> If necessary, transformation of the variables has been applied to meet the model assumptions. When of interest, pairwise Tukey *post hoc* comparisons have been carried out.

Model-based parametric bootstrap was performed to calculate model predictions, using the bootMer function from LME4.<sup>24</sup> One thousand simulations were run for each model. In each simulation, new values of the random effects and the residual errors were generated.

## III. RESULTS

### A. Retaining information by speaker and listener

No significant relationship was found between the exam score and gender ( $p > 0.1$ ). The order of the roles (first listener, then speaker, or the other way around) had also no significant influence on the exam score ( $p > 0.1$ ), and, interestingly, neither had the type of background noise ( $p > 0.1$ ).

The finally retained model is given in Table I. Participants scored significantly better on questions about lectures they had read out loud (speaker's role) compared to those they had listened to (listener's role) ( $p < 0.00001$ ). Participants' scores also differed depending on their interest in the topic ( $p < 0.00001$ ). Tukey *post hoc* comparison (Table I) shows that scores were especially higher for topics the participants were interested in.

### B. Secondary task listener

Performance on the secondary visual task was consistently very high, with in total 88.8% observations scoring perfectly (57.2%) or making one mistake (31.6%). Because the performance varied so little, no mixed models were calculated for this variable, instead duration of the task was used as outcome variable.

No significant relationship was found between task duration and gender ( $p > 0.1$ ), and also not between task duration and order of the roles (first listener, then speaker, or the other way around) ( $p > 0.05$ ).

The final model is summarized in Table II. No significant interaction effect between background noise and the trial order of the visual tasks (first, second, or third) was found ( $p > 0.1$ ).

The type of background noise had a clearly significant influence on the duration of the task ( $p < 0.0001$ ). Figure 2 shows the predicted tasks duration as a function of background noise and trial order of the visual task. *Post hoc* analyses (Table II) show that reaction time was consequently longer for fluctuating city street noise compared to the other background noise conditions.

It was expected that the reaction time would decrease as the participant's gained experience with the visual task ( $p < 0.01$ ). Indeed, the shortest reaction time was observed for the last trial (Table II). Table II also clearly shows that the effect of trial order on task duration is less pronounced than the effect of background noise.



TABLE I. Overview of final mixed model for the dependent variable exam score: random effects (Participant and Lecture) with residual error, fixed effects (Role and Interest), and results of single term deletion. For the random effects, the variance and corresponding standard deviation (Std. Dev.) are given. For the fixed effects, the estimated coefficients are shown (Estimate), together with their standard error (Std. Error) and t-value. For the single term deletion, degrees of freedom (DF), Akaike information criterion (AIC), likelihood ratio test (LRT), and corresponding  $p$ -value (Pr(Chi)) are tabulated, \*\*\* is  $p < 0.0001$ . Finally significant Tukey *post hoc* comparisons are shown.

Random effects:				
Variable	Variance		Std. dev.	
Participant	0.7927		0.8903	
Lecture	0.5994		0.7742	
Residual	2.4061		1.5512	
Fixed effects:				
	Estimate	Std. error	t value	
(Intercept)	3.0596	0.3305	9.257	
RoleSpeaker	0.8283	0.1417	5.844	
InterestNo	−1.0948	0.1914	−5.721	
InterestNeutral	−0.7667	0.2025	−3.786	
Single term deletion:				
	Df	AIC	LRT	Pr(Chi)
None		1892.0		
Role	1	1923.0	32.974	9.341e-09***
Interest	2	1921.4	33.391	5.613e-08***
Significant differences in exam score for different levels of variable Interest (Tukey <i>post hoc</i> testing):				
			p-value	
Interested	versus	Not interested	<0.00001	
Interested	versus	Neutral	<0.001	

### C. Hesitation phenomena speaker

No significant relationship was found between the duration of disfluencies and the order of the role (first speaker, then listener, or the other way around) ( $p > 0.1$ ), interest in the topic ( $p > 0.1$ ), or gender ( $p > 0.05$ ).

In the final model, duration of the disfluencies has been modeled as a function of background noise (fixed effect), participant (random effect), and lecture (random effect). The model is summarized in Table III.

Background noise has a statistically significant influence on disfluency duration ( $p < 0.001$ ). *Post hoc* comparison shows that averaged duration of the disfluencies is especially shorter in fluctuating city street noise compared to, respectively, highway noise and quiet, see Fig. 3 and Table III. For disfluency duration in city street noise compared to multi-talker babble, no statistically significant difference was found ( $p > 0.1$ ).

### D. Information transfer

To assess the transfer of information, the difference is made between the exam score of the speaker, and the score

TABLE II. Overview of final mixed model for the dependent variable secondary task duration: random (Participant and Lecture) effects with residual error, fixed effects (Background noise condition and Test sequence), and results of single term deletion, \*\*\* is  $p < 0.0001$ , \*\* is  $p < 0.01$ . For full explanation on table structure see Table I.

Random effects:				
Variable	Variance	Std. dev.		
Participant	7.102e-03	0.084273		
Lecture	2.915e-06	0.001707		
Residual	7.810e-03	0.088374		
Fixed effects:				
	Estimate	Std. Error	t value	
(Intercept)	1.014618	0.014416	70.38	
Noisehighway	−0.031480	0.010935	−2.88	
Noisebabble	−0.031349	0.010540	−2.97	
Noisequiet	−0.051386	0.010339	−4.97	
Sequence2	0.006168	0.008551	0.72	
Sequence3	−0.021514	0.008543	−2.52	
Single term deletion:				
	Df	AIC	LRT	Pr(Chi)
none		−1157.0		
Noise	3	−1138.5	24.506	1.958e-05***
Sequence	2	−1149.5	11.488	0.003201**
Significant differences in secondary task duration for different levels of variable Noise (Tukey <i>post hoc</i> testing):				
				<i>p</i> -value
City	versus	Quiet	<0.001	
City	versus	Babble	<0.05	
City	versus	Highway	<0.05	
Significant differences in secondary task duration for different levels of variable Test sequence (Tukey <i>post hoc</i> testing):				
				<i>p</i> -value
3	versus	1	<0.01	
3	versus	2	<0.01	

of corresponding listener. The reasoning behind this is that similarity in score between speaker and listener represent the overall difficulty of the topic. The difference in exam score is partially attributable to information lost during the communication process. The hypothesis is that if the communication process is hard because of more hesitation from the speaker's side, more information might be lost.

A significantly negative correlation (Pearson's  $\rho = -0.244$ ;  $p < 0.0001$ ) between the duration of disfluencies (logarithmic transformation) and the score of the exam is found, suggesting that indeed more information is lost in the communication process when the duration of the disfluencies increases. Obviously the duration of disfluencies can only account for a limited part of the variation in obtained score, as individually different characteristics between speaker and listener will also influence the difference in exam score.

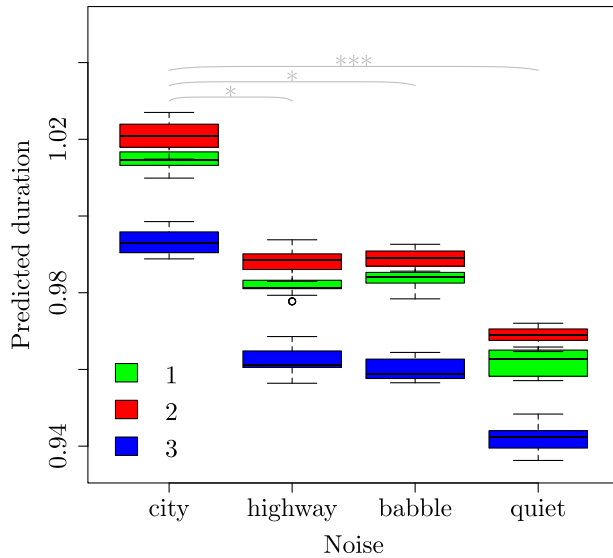


FIG. 2. (Color online) Box-and-whisker plot of median of the predicted visual task duration (log-transformed) as a function of background noise conditions (on the  $x$  axis) and of trial of the visual tasks (first, second, or third) per lecture. The branches represent pair-wise significant differences between background noise conditions as found by Tukey *post hoc* testing \*\*\* is  $p < 0.0001$ , \* is  $p < 0.05$ . For the box-and-whisker plot, the bottom and top of the box are the first and third quartiles, the band inside the box is the median. The whiskers extend to the most extreme data point which lies no more than 1.5 times the interquartile range above the upper, or below the lower quartile. The circles represent outliers, i.e., data points located outside the whiskers.

#### IV. DISCUSSION

This paper assesses the effects of background noise on presenting and perceiving information. The most distinct result is the effect of background noise on the secondary task listeners had to perform; the response time clearly differs depending on the type of background noise. The background noise fragment with city street noise appeared to be the most disturbing. This noise fragment is very salient, with relatively long quiet periods suddenly interrupted by distinct sounds clearly emerging from the background.<sup>26</sup>

Variation in speech perception on itself is insufficient to explain the observed effects of background noise, because then better scores are expected with more fluctuating maskers due to release from masking.<sup>10</sup> The obtained results are more in line with the known disruptive effects of distinctive temporal-spectral variations on attention and focusing.<sup>11,12</sup>

Unlike what would be expected based on existing literature, fluctuating city street sound was more disturbing than the multitalker babble. In generally, speech is expected to have a larger effect compared to non-speech maskers.<sup>27</sup> A possible explanation is the large numbers of talkers in the multitalker babble, about 20 people talking at a cocktail party. The effect of competing talkers becomes less prominent as the number of talkers increases, around 12 the masking effects saturates because the talkers in the babble are masking each other.<sup>9</sup>

Masking will not only affect speech intelligibility, it also implies that particular variations in frequency and intensity will to a certain extent be averaged out in the amalgam

TABLE III. Overview of final mixed model for the independent variable duration of disfluencies: random effects (Participant and Lecture), residual error, fixed effects (Background noise condition), and single term deletion, \*\* is  $p < 0.01$ . For full explanation on table structure see Table I.

Random effects:				
Variable	Variance	Std. dev.		
Participant	0.0113715	0.1066		
Lecture	0.0008352	0.0289		
Residual	0.0120422	0.1097		
Fixed effects:				
	Estimate	Std. error	t value	
(Intercept)	−0.38258	0.02232	−17.141	
Noisehighway	0.06721	0.02033	3.306	
Noisebabble	0.02498	0.02038	1.226	
Noisequiet	0.05293	0.02018	2.623	
Single term deletion:				
	Df	AIC	LRT	Pr(Chi)
none		−268.09		
Noise	3	−261.18	12.91	0.004836**
Significant differences in duration of disfluencies for different levels of variable Noise (Tukey <i>post hoc</i> testing):				
				<i>p</i> -value
City	versus	Highway	<0.001	
City	versus	Quiet	<0.05	

of voices. Therefore, the novelty detection system of human listeners might be triggered less, whereas it is more triggered by the non-speech eventful sounds in the fluctuating city street noise fragment. This detection system is part of

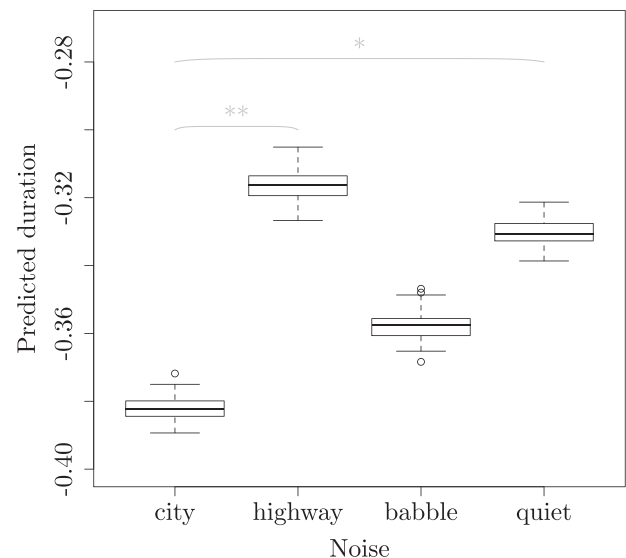


FIG. 3. Box-and-whisker plot of median of the predicted hesitation duration (log-transformed) as a function of background noise condition. The branches represent pair-wise significant differences between the background noise conditions as found by Tukey *post hoc* testing, \*\* is  $p < 0.01$ , \* is  $p < 0.05$ . For the general structure of box-and-whisker plots, see Fig. 2 for the full explanation.

bottom-up mechanisms of selective attention and constantly monitors the acoustical environment for changes in frequency, intensity, duration, and spatial location.<sup>28</sup> These findings do warrant us to be vigilant when it comes to common yet eventful sounds in a teaching context. Most work on the effects of environmental noise focuses on sounds with high sound pressure level and/or (intelligible and meaningful) speech. This work shows that also moderate-level non-speech sounds, such as one car passing by, might affect information processing.

Both performance and effort have been assessed to address the effects of background noise on information processing. The task performance (exam results) appeared to be relatively unaffected, possibly because the noise levels were only moderately high. Noise fragments were played at about 70 dB LAeq measured over the whole duration of each noise fragment. This level is in accordance to background noise levels reported from Danish, British, and Swedish schools.<sup>16</sup> However, it is somewhat lower compared to levels cited by Ref. 27, reviewing the effects of background noise on task performance in general. In addition, proximity and visibility of the speaker gave good visual information, facilitating speech perception and hence reducing noise interference.<sup>29</sup>

Listening effort is assessed by a secondary visual task to be performed during noise exposure. The time needed for the task appeared to be a more informative parameter than performance, which is in accordance to previous findings studying task performance in noise inside high speed trains.<sup>30</sup>

The current test design deviates from a classic dual-task paradigm. In a classic dual-task paradigm, the primary and the secondary task would also have been performed separately in single task conditions.<sup>5</sup> Practical constraints did not allow us to include these single task conditions, the current protocol was already demanding for the participants. Therefore, it cannot be stated unambiguously that performance on the secondary task varied because of the demands of the primary task. It is also theoretically possible that the secondary task has influenced performance on the primary task. However, participants appear to be quite capable to focus on a primary task when instructed to.<sup>5</sup> In addition, the secondary task had to be completed only on a few occasions during the lectures, no more than five times.

The current design does resemble a realistic learning context where different tasks have to be performed at the same time. Seeing no effect of background noise on the primary task whereas the secondary task is altered, suggests that in certain noise conditions cognitive resources and cognitive load had to be allocated from the secondary to the primary task to maintain performance on the primary task. This finding is a strong argument to invest in test protocols including a close-to-reality variety of sufficiently complex tasks. Otherwise, the effects of background noise in real learning environments is at risk to be incorrectly predicted.<sup>18,31</sup>

For the speaker, hesitation phenomena observed in speech were significantly influenced by the background noise. Interestingly, the most difficult noise conditions from the listener's perspective, i.e., fluctuating city street noise,

appeared to reduce the occurrence of hesitation phenomena. This suggests that speakers focused more on their speech in this particular condition; availability of attentional resources has indeed been related to occurrence of disfluencies.<sup>32</sup>

A higher focus on speech in more difficult background noise conditions might have been triggered by three different processes. First, it is plausible that, in general, reading the text out loud was a relatively easy task for the speakers. It is known that for easier tasks, more challenging background noise might actually increase performance because it activates additional focus and attention that is not required in less challenging background noise conditions.<sup>33</sup> Second, the speaker could have been aware that the re-occurring salient events in the fluctuating city street noise would incorrectly trigger their novelty detection system. In order not to fall into the trap of attention shift to irrelevant stimuli, the speaker could have put extra focus on the task at hand, resulting in more fluent speech. Third, the speaker could have identified the fluctuating noise as the most difficult condition from the listener's perspective. Paying more attention to speech (fluency) would then be done to ensure optimal communication and information transfer.

All explanations for reduced hesitation phenomena suggest that the task for the speakers was relatively easy, as they could, unplanned or deliberately, adapt their speech production in more difficult listening conditions. If their tasks would have been more demanding, the necessary allocation of cognitive resources to improve speech production might not have been possible. In this, the current design might underestimate the effects on speakers as in most teaching conditions information is given via spontaneous speech and not read out loud. Then again, contrary to real teaching situations, speakers had no or little prior knowledge of the information to be presented. They had to acquire the information while reading to be able to complete the exam afterwards. This added extra challenges for the speaker that would normally be less present in real teaching conditions.

Understanding the effect of background noise in ecologically valid teaching situations is the key element of this study. As a consequence, the included tasks had to be made representative for presenting and processing information in real-world conditions. The disadvantage of this approach is that such complex tasks are process impure; instead of addressing one particular cognitive process, a whole set of processes is implied to fulfill the tasks at hand. This means that the final results, i.e., variation in performance of both speaker and listener depending on the background noise, cannot be unambiguously pin-pointed to the effect of background noise on particular cognitive process, such as serial rehearsal or episodic memory. Contrary, process-specific tasks are not suitable to predict the effect of noise in real-world conditions where the outcome is determined by the interaction of a whole set of cognitive processes.<sup>18</sup> As this project aims to understand the effects of noise on real-world information transfer, using complex ecologically valid tasks is therefore crucial.

A second caveat when working with ecologically valid designs is increase in test variability. In this design, participants had been invited in duos, and each participant both

took the role of listener and speaker. This means that for each listener information was presented by a different speaker, which is a substantial source of variability compared to research designs where uniform stimuli are presented to all listeners. Increase in test result variability implies that effects of background noise needed to be rather robust to lead to statistically significant results.

In general, not using standardized speech perception/production tests makes the question of result reproducibility more pertinent. Within the constraints of an ecologically valid design, fundamental principles of reproducible test design have been respected by carefully controlling noise exposure and experimental conditions for both speaker and listener, as well as establishing an overall protocol that was very strictly followed for all tests.

One could argue that in this particular study part of the challenges related to ecologically valid design arise from the decision of including speaker and listener in the same setup. Having a design where speaker and listener are tested on separate occasions could have substantially simplified testing and make interpretation more straightforward. However, such a separate approach, although commonly applied, is incapable of capturing effects very pertinent in real-life communication in noise. First, for the speaker having an actual listener to talk to is crucial, as it has been shown that speakers responded differently to the background noise when they were asked to read instruction out loud alone, or when communicating them to another participant.<sup>19</sup> These changes were not only quantitative, but also qualitative.<sup>19</sup>

Furthermore, for the listener, presenting prerecorded material in background noise instead of a live speaker is cumbersome. While it is quite well-understood how the overall speech levels are adapted to overall background noise,<sup>34</sup> changes in frequency content and temporal characteristics are much more challenging to apply, left alone the effect of short-time fluctuations in background noise. Therefore, adapting speech material recorded in silence to speech produced in noise will to a certain extent remain artificial and it is unlikely that all (acoustical) changes are accurately included. Alternatively, having one particular speaker for all fragments in noise risks that particularities of one speaker will determine the result, and therefore that findings cannot be generalized.

Another important factor in real-life communication is visual information, including lip-reading.<sup>5</sup> This is included in the current design. The fact that in most of the cases speaker and listener knew each other beforehand is also close to a realistic teaching situation. Finally, unlike general one-to-one communication settings, interaction between speaker and listener was limited in this setup, which is also more realistic to general teaching conditions where information transfer is largely steered by the teacher.

The current study in particular has pointed towards interesting findings with respect to interaction between speaker and listener in fluctuating background noise. These results suggest that naive speakers, unaware of the formal acoustical characteristics of the background noise, might have been systematically capable of identifying the listening condition that was indeed most difficult for the listeners, i.e.,

fluctuating city street noise. These results suggest that (controlling) hesitation phenomena should be added to potential speech parameters changing in background noise, in addition to known changes in level, spectral, and temporal speech characteristics.<sup>14</sup> It is unclear whether speakers would have responded similarly in the absence of a listener, and it is clear that controlling/introducing hesitation phenomena in prerecorded speech would be highly unlikely to be done in a realistic way. These findings therefore certainly advocate for combined speaker–listener study design when assessing the influence of background noise on information processing.

Further research will focus on protocols that allow more interaction between speaker and listener, including interruptions and questions. Conversation analysis<sup>35</sup> could offer great potential to study the effects of background noise on information transfer. Visual interaction and gestures would also be worthwhile to be taken into account. In addition, insight in the listener's response to speech and background noise will be substantially increased by capturing the acoustical signals at the listener's ear with binaural in-ear microphones, as is done sometimes done for characterization of environmental sound and soundscapes.<sup>36</sup> This will allow us to quantify acoustical parameters, such as signal-to-noise ratio, most relevant for the listener, i.e., the input directly received by the listener.

Developing standardized test protocols can be considered as one of the major challenges in ecologically valid research. Efforts are being made to provide more standardized protocols that do include concepts such as effortful listening.<sup>3</sup> To really unravel the underlying mechanisms of auditory processing in noise, a combined approach with more ecologically valid and more standardized test could be envisaged, preferably including behavioral testing, electrophysiological measurements, and self-reports.

## V. CONCLUSION

The influence of background noise on auditory processing and information transfer has been assessed. For the listener, the background noise that is the most eventful and salient appears to be the most difficult. Especially the secondary visual tasks take longer in this condition. Interestingly, this is the same condition where speakers appear to be most focused, potentially to accommodate for the challenges of this background noise condition so that adequate information transfer is maintained. The applied protocol shows the potential of an ecologically valid test design to understand real-life effects of background noise based on experimental research.

<sup>1</sup>C. Clark and P. Sörqvist, "A 3 year update on the influence of noise on performance and behavior," *Noise Health* **14**, 292–296 (2012).

<sup>2</sup>R. Ljung, P. Sörqvist, A. Kjellberg, and A.-M. Green, "Poor listening conditions impair memory for intelligible lectures: Implications for acoustic classroom standards," *Build. Acoust.* **16**, 257–265 (2009).

<sup>3</sup>M. K. Pichora-Fuller, S. E. Kramer, M. A. Eckert, B. Edwards, B. W. Hornsby, L. E. Humes, U. Lemke, T. Lunner, M. Matthen, C. L. Mackersie, G. Naylor, N. A. Phillips, M. Richter, M. Rudner, M. S. Sommers, K. L. Tremblay, and A. Wingfield, "Hearing impairment and cognitive energy: The framework for understanding effortful listening (fuel)," *Ear Hear.* **37**, 5S–27S (2016).



- <sup>4</sup>M. Rudner, "Cognitive spare capacity as an index of listening effort," *Ear Hear.* **37**, 69S–76S (2016).
- <sup>5</sup>J.-P. Gagné, J. Besser, and U. Lemke, "Behavioral assessment of listening effort using a dual-task paradigm: A review," *Trends Hear.* **21**, 1–25 (2017).
- <sup>6</sup>P. A. Tun, A. Wingfield, and E. A. Stine, "Speech-processing capacity in young and older adults: A dual-task study," *Psychol. Aging* **6**, 3–9 (1991).
- <sup>7</sup>B. Rakerd, P. Seitz, and M. Whearty, "Assessing the cognitive demands of speech listening for people with hearing losses," *Ear Hear.* **17**, 97–106 (1996).
- <sup>8</sup>S. Clarke and P. Morosan, *The Human Auditory Cortex* (Springer, New York, 2012), pp. 11–38.
- <sup>9</sup>E. H. N. Ng, M. Rudner, T. Lunner, and J. Rönnberg, "Noise reduction improves memory for target language speech in competing native but not foreign language speech," *Ear Hear.* **36**, 82–91 (2015).
- <sup>10</sup>K. S. Rhebergen, N. J. Versfeld, and W. A. Dreschler, "Release from informational masking by time reversal of native and non-native interfering speech," *J. Acoust. Soc. Am.* **118**, 1274–1277 (2005).
- <sup>11</sup>S. J. Schlittmeier, T. Weißgerber, S. Kerber, H. Fastl, and J. Hellbrück, "Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength," *Atten. Percept. Psychophys.* **74**, 194–203 (2012).
- <sup>12</sup>S. P. Banbury, W. J. Macken, S. Tremblay, and D. M. Jones, "Auditory distraction and short-term memory: Phenomena and practical implications," *Hum. Factors* **43**, 12–29 (2001).
- <sup>13</sup>H. Jahncke, S. Hygge, N. Halin, A. M. Green, and K. Dimberg, "Open-plan office noise: Cognitive performance and restoration," *J. Environ. Psychol.* **31**, 373–382 (2011).
- <sup>14</sup>C. Hotchkin and S. Parks, "The Lombard effect and other noise-induced vocal modifications: Insight from mammalian communication systems," *Biol. Rev.* **88**, 809–824 (2013).
- <sup>15</sup>V. Aubanel and M. Cooke, "Strategies adopted by talkers faced with fluctuating and competing-speech maskers," *J. Acoust. Soc. Am.* **134**, 2884–2894 (2013).
- <sup>16</sup>J. Kristiansen, S. P. Lund, R. Persson, H. Shibuya, P. M. Nielsen, and M. Scholz, "A study of classroom acoustics and school teachers' noise exposure, voice load and speaking time during teaching, and the effects on vocal and mental fatigue development," *Int. Arch. Occup. Environ. Health* **87**, 851–860 (2014).
- <sup>17</sup>C. Fehrer and C. Fry, "Hesitation phenomena in the language production of bilingual speakers: The role of working memory," *Folia Linguist.* **41**, 37–72 (2007).
- <sup>18</sup>P. Sörqvist, "On interpretation and task selection in studies on the effects of noise on cognitive performance," *Front. Psychol.* **5**, 1–4 (2014).
- <sup>19</sup>M. Garnier, N. Henrich, and D. Dubois, "Influence of sound immersion and communicative interaction on the Lombard effect," *J. Speech Lang. Hear. Res.* **53**, 588–608 (2010).
- <sup>20</sup>ISO 532-1:2017, "Acoustics—Methods for calculating loudness—Part 1: Zwicker method," 1st ed. (Technical Committee, ISO/TC 43 Acoustics, 2017).
- <sup>21</sup>M. Florentine, *Loudness* (Springer, New York, 2011), Chap. 1.
- <sup>22</sup>S. Degeest, H. Keppler, and P. Corthals, "The effect of age on listening effort," *J. Speech Lang. Hear. Res.* **58**, 1592–1600 (2015).
- <sup>23</sup>George Neumann GmbH, *Operating instructions for Neumann condenser microphones of the f e t 80-series, KM 83 i, KM 84 i, KM 85 i, KM 86 i, KM 88 i, KMS 84 i*, D-1000 Berlin 61, Charlottenstrasse 3 (1984).
- <sup>24</sup>D. Bates, M. Maechler, B. Bolker, S. Walker, R. H. B. Christensen, H. Singmann, B. Dai, G. Grothendieck, P. Green, and M. B. Bolker, "Package 'lme4,'" in *R Package Version 1.1-10* (2016).
- <sup>25</sup>M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, *Applied Linear Statistical Models* (McGraw-Hill, New York, 2005).
- <sup>26</sup>B. De Coensel, D. Botteldooren, T. De Muer, B. Berglund, M. Nilsson, and P. Lercher, "A model for the perception of environmental sound based on notice-events," *J. Acoust. Soc. Am.* **126**, 656–665 (2009).
- <sup>27</sup>J. L. Szalma and P. A. Hancock, "Noise effects on human performance: A meta-analytic synthesis," *Psychol. Bull.* **137**, 682–707 (2011).
- <sup>28</sup>D. Oldoni, B. De Coensel, M. Boes, M. Rademaker, B. De Baets, T. Van Renterghem, and D. Botteldooren, "A computational model for auditory attention of use in soundscape research," *J. Acoust. Soc. Am.* **134**, 852–861 (2013).
- <sup>29</sup>L. Girin, J.-L. Schwartz, and G. Feng, "Audio-visual enhancement of speech in noise," *J. Acoust. Soc. Am.* **109**, 3007–3020 (2001).
- <sup>30</sup>W. Wei, A. Bockstael, B. De Coensel, and D. Botteldooren, "Interference of speech and interior noise of Chinese high-speed trains with task performance," *Acta Acust. united Acust.* **98**, 790–799 (2012).
- <sup>31</sup>C. P. Beaman, "Auditory distraction from low-intensity noise: A review of the consequences for learning and workplace environments," *Appl. Cognit. Psychol.* **19**, 1041–1064 (2005).
- <sup>32</sup>C. C. Oomen and A. Postma, "Effects of divided attention on the production of filled pauses and repetitions," *J. Speech Lang. Hear. Res.* **44**, 997–1004 (2001).
- <sup>33</sup>N. Halin, J. E. Marsh, A. Hellman, I. Hellström, and P. Sörqvist, "A shield against distraction," *J. Appl. Res. Mem. Cognit.* **3**, 31–36 (2014).
- <sup>34</sup>P. Corthals, "Sound pressure level of running speech: Percentile level statistics and equivalent continuous sound level," *Folia Phoniatr. Logopaedica* **56**, 170–181 (2004).
- <sup>35</sup>J. Sidnell and T. Stivers, *The Handbook of Conversation Analysis* (Wiley, New York, 2012), Vol. 121.
- <sup>36</sup>M. Rychtáriková and G. Vermeir, "Soundscape categorization on the basis of objective acoustical parameters," *Appl. Acoust.* **74**, 240–247 (2013).